Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2007 Proceedings

Americas Conference on Information Systems (AMCIS)

December 2007

Data Quality Begins with the Business Model

Mary-Ann Robbert
Bentley College

Donna Fletcher
Bentley College

DATA QUALITY BEGINS WITH THE BUSINESS MODEL

Mary Ann Robbert
Bentley College
mrobbert@bentley.edu

Donna Brown Fletcher
Bentley College
dfletcher@bentley.edu

Abstract

This paper examines the effect of the order of a system redesign process on the quality of the information. A process that focuses on understanding the business model and the data prior to making technical decisions can produce information of higher quality, while continual use of patches and technology solutions can reduce the system's data quality. The study of a global financial institution is used to exemplify the issues.

Keywords

System redesign, data quality, information quality

Introduction

Organizations are not static and neither are their systems. Systems continually evolve and this evolution can take place in many different forms: systems integration, new products added, hardware or software upgrades, etc. In any case the system must be redesigned or rebuilt. Simplistically, this can be viewed in two primary ways, i.e. systems can be redesigned or modified by using a top down or bottom up approach. Proponents of each method tout the benefits of their preference on the system design or program (Hakos 1998, McCorquodale 2003). In this paper we will focus on the effect of system redesign on information quality though most of the concepts can also be applied to system design..

We allocate four different components for the system redesign: technology planning, application planning, data comprehension, and business requirements (based on those proposed by panel, Yonke 2006). Each component has a position in the sequence of design events and a weight based on time and cost. The choice of sequence number and the weight given to each level influences the system's information quality as well as the design implementation.

Our analysis is based on a large financial institution with a complex reporting system that we have studied over the past two years. The methodology we advocate is not currently used or promoted by the company. While we isolate and dissect the concepts within component parts of the system, the development of the system continues through our analysis. This research studies the effect of the model redesign structure on the quality of the data as the system functions.

After current research is reviewed, the next section of the paper discusses the research model. This is followed by a discussion of the realities that come between the theory and the implementation.

Related Research

System Redesign

Large global organizations must deal with complex technical and managerial systems that continually change. Modeling these systems and revisions is a challenge. Bubenko noted in the eighties that there were hundreds if not thousands of information systems development methods (Bubenko 1986). Fitzgerald tried to ascertain whether a classification of systems development based on the use of specific formalized methodologies is useful in differentiating the development process in organizations. He concludes that this is not the case, noting that the solution to systems development problems in terms of increased control and the more widespread adoption of formalized development methodologies has not been proven (Fitzgerald 1998)**.** We are not aware of any studies of the effects of the various methodologies on the quality of the information. However, though the top down/ bottom up approach have not been explored in the literature, it does frequently ground the discussion in practice.

In response to the need for a more flexible approach, a "framework" perspective has been used for evaluation of a system in context to its business environment. Woo et al (2006) review system development methodologies and propose a framework to track the implementation of the system development methodologies. They conclude that a robust framework can significantly save time and money, plus reduce the number of problems (Woo 2006). Further, object models have been utilized in modeling system design. These models are based on an abstract system model, using concepts and relationships (Monarchi 1992). Object models are frequently use-case driven, and have become popular with the introduction of UML. However, object models are not considered in this paper, since they represent a totally different paradigm. The approach we advocate, similar to that proposed by McCorquodale (2003), views the system within the context of the business model which the system supports.

Data quality

Today's enterprise realizes the need for information quality. From a $1 million cost increase for regrinding of lenses because of data errors (Wang 1998) to the loss of lives from the Vincennes shooting down of an Iranian passenger plane (Fisher 2001), many examples of errors caused by poor data have been recorded. Over the past two decades researchers such as Richard Wang, Lee Yang, Tom Redman and many others have analyzed the causes of poor data and proposed solutions (Wand 1996, Wang 1998, Lee 2002, Redman 1998).

Strong, Lee and Wang (1997) categorize information quality into four major aspects with associated quality dimensions as follows:

Category	***Dimension***
Intrinsic IQ	Accuracy, Objectivity, Believability, Reputation
Accessibility IQ	Accessibility, Security
Contextual IQ	Relevancy, Value-Added, Timeliness, Amount of Information
Representational IQ	Interpretability, Ease of Understanding, Concise Representation, Consistent Representation

The data quality dimensions that are of most concern in our research study are the intrinsic and contextual data quality dimensions. As noted in Fletcher, et. al. (2005) systematic errors in information production can lead to lost information, resulting in errors of correctness, completeness and relevancy. Systematic errors during production are especially important because they affect the entire system and arguably, these errors are more common as the system evolves during the redesign process.

The problem of poor data quality has been analyzed and measured from a technical perspective (Liu 2006, DeAmicis 2006). De Amicis et al specifically examined financial information, especially registry data, and designed methods to measure the qualitative and quantitative quality of the data (DeAmicis 2006)

Delez and Hostettler show the problem of poor data quality can be approached as a business problem rather than a technical problem noting that inadequate information quality can lead to the collapse of enterprise-wide IT projects. Adequate information quality can secure a timely and accurate basis for decision-making, supply chain efficiency, IT business case realization and reduction in transaction costs. (Delez 2006). Further, Raneses et al show how data quality is linked to failed business processes and therefore poses operational risks (Raneses 2006). It is agreed that information quality relies on the successful leverage of business knowledge.

In financial institutions there is not only a need for high quality data there is a regulatory requirement. In order for banks to qualify to use the Internal Ratings Based Approach to calculating regulatory and economic capital, The Basel II Accord requires validation of the banks internal models (Basel II 2005, Ong 2005). Without data or with poor quality data, models cannot be validated.

Model

The order in which a system is redesigned or modified will affect the quality of the information produced by the system. We assume the goal of the system is to produce the highest quality of information. For this study we simplified the model to see if the weight and placement of the components could affect the information quality. We first considered a top down model of system redesign comprised of four steps: plan technology, plan applications, understand data, and understand the business model (Yonke 2006), ordered as shown in Figure 1. Top down in this case refers to the direction in which the design steps are completed rather than the systems design definition of proceeding from globally accessible to available subcomponents (Crespi 2005) In Figures 1, 2 and 3, the width of the layer signifies the weight (cost or time spent) while the arrow indicates the order of completion. Thus in the top down model, the organization allocates the most resources and time to technology which is studied and implemented first, an occurrence in many companies.

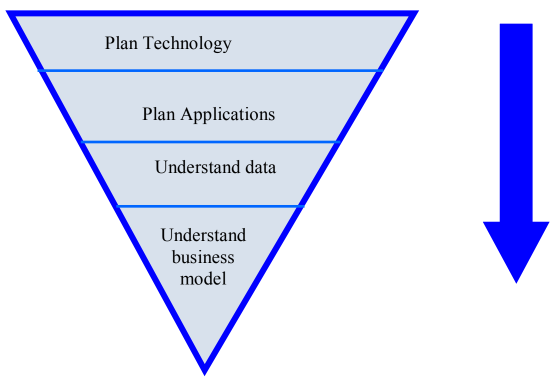

Figure 1
Top down model

The existing system bounds ideas for the new system. The redesign focuses on technical solutions to errors in the current system. After the technology has been examined and plans for the new system modeled, required applications are planned. These applications will then fit the technology that has already been implemented or at least selected. It is not until this third level that the data is examined. The data is tested with the primary focus on the dimensions of accuracy and availability. Finally the business model is examined to verify that the system performs as specified. Frequently problems occur in this top-down approach, i.e. either the business model has to be modified so that it will fit the system, the planned system is patched or the business model and the system exist out of synch.

Conceptually, this design approach represents an upside down triangle and therefore requires a dedicated person to hold up the triangle so that it doesn't fall. Or, the entire IT group supports the structure and maintains the stability of the system. This is very costly in time and effort and can still result in the collapse of the structure. In this case data is constantly cleaned and processes are continually redesigned to maintain the required degree of data quality for the subsequent reports.

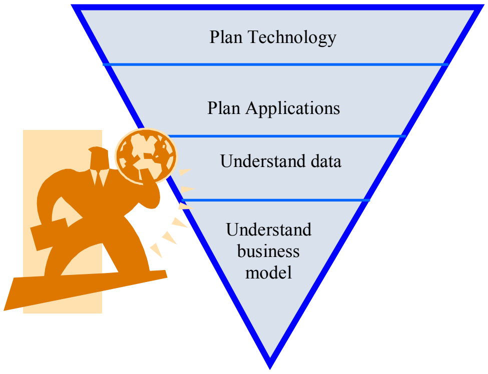

Figure 2
System upheld by technician

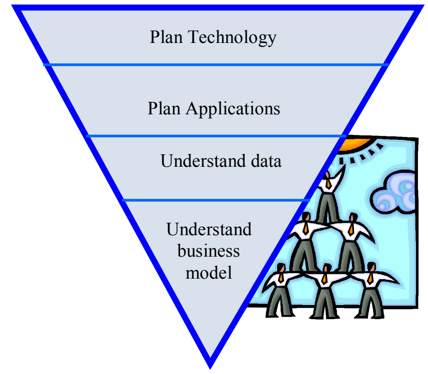

Figure 3
System upheld by technical team

Alternatively, the bottom up model recognizes the business model as the basis or the foundation for the development of the entire system (see Figure 4). Now the system designers are forced to examine the business first to understand the business model. Again note we refer to bottom up as weight and direction as opposed to specifying components first then linking them together. With the firm foundation of a well defined business model, data is then examined. Delez and Hostettler designed a grid and demonstrated that establishing a solid foundation with business definitions and data processes as the most effective way to achieve self-sustained information quality (Delez 2006). Examining the data before any applications are built or any other technology is planned prioritizes the data quality problems. The applications can be designed to address any quality issues that have been already identified. Finally the required technology will be modeled with data quality needs in mind. This approach can be applied whether current technology will remain in place or a new system is being built.

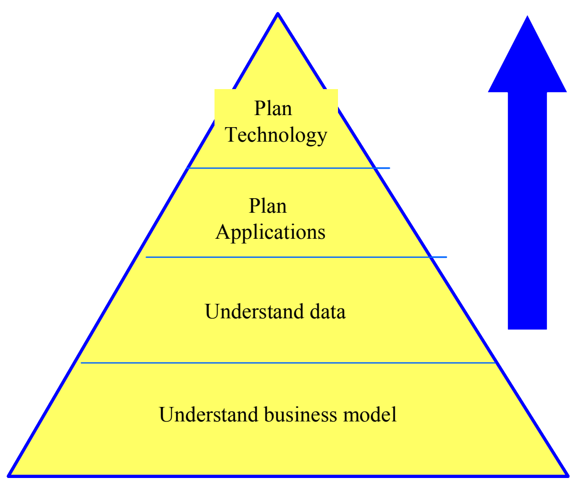

Figure 3
Bottom up model

Beginning system redesign based on the business model results in a higher probability that the technology will then be constructed to support the business model. Analyzing the data at this stage may alleviate more complicated data quality issues further along in the system redesign. As already stated, the business model defines the intrinsic and contextual data quality dimensions which need to be factored into the system redesign**.**

The remainder of the paper provides detail on our bottom up approach and the benefits in using this approach in determining data quality. We next discuss the business model used for calculating credit risk exposure and the data required for this model.

Business model and data requirements

The business model of concern is the determination of regulatory capital for credit risk exposure. Basel II requires calculation of probability of default (PD), loss given default (LGD) and exposure at default (EAD) parameters, which when multiplied together result in expected loss (EL). One measure of credit risk exposure that impacts EAD is Pre-Settlement Exposure (PSE) associated with derivative contracts. PSE is the exposure that occurs when a counterparty becomes unable to settle its contractual obligations (contracts) prior to settlement such that the bank therefore needs to "replace" these contracts in the market.

The data needed for the determination of PSE is dependent on whether a transactions approach or a portfolio approach is used. The portfolio approach is preferred, as it allows for account netting and its effect, and represents a time varying risk assessment over the life of the portfolio, with a peak risk exposure occurring at a particular point in time. However, whereas the transactions approach uses general market rates and volatilities and standard contracts, the portfolio approach uses actual market rates, volatilities and correlations for the actual contracts in the portfolio. Thus, the data requirements for the portfolio approach to PSE measurement are more onerous.

Fletcher, et al 2005 note in the initial phase of this research that an unacceptably high number of transactions could not be leveraged by the Credit Systems for PSE calculations using the preferred portfolio based Monte Carlo simulation methodology for the reasons listed below. Consequently, the institution must add the transaction based PSE for these incomplete contracts to the portfolio based result to arrive at the overall credit risk exposure measurement:

- ✓ Unsupported products (products for which pricing models and market data simulation do not exist in the PSE Server).
- ✓ Source system contains flawed handoff logic to the Credit Systems for transaction and market data.
- ✓ No comprehensive transaction and market data model(s) to describe the data required by credit necessary to calculate exposure accurately.
- ✓ Credit Sytems infrastructure has organizational design issues. The lines of reporting and responsibility within the Credit Systems have been blurred as the organization has made changes to support the new regulatory needs.
- ✓ Missing Service Level Agreements (SLAs) for market data
- ✓ Correct data validation and error trapping doesn't occur at the earliest point possible within the Credit Systems.

Yet, simply addressing these problems is not sufficient. Understanding the data needs of the business model also requires a determination of which dimensions of quality are important for each user group along the path. In a large system many groups rely on the data for different functions. These different functions will then determine the quality required for their specific needs and the corresponding dimensions that must be analyzed. Knowledge of the data quality needs of the user groups and how these groups address the data quality issues is important to determining the overall of data quality of the system.

Plan applications

Consequently, it is necessary to determine what user groups need in terms of data and how they interface with the data. Applications must support both the users and the business model. This layer defines what applications will do to manage the data and provide information to performing business functions.

Spewak defines an applications architecture that would be constructed here. His architecture includes: list candidate applications, including those that can improve the business; define applications and ensuring they are understandable to users and support business functions; relate applications to functions understanding the related business functions and analyzing the impact to current applications, relating applications to existing system (Spewak, 1992). Using the bottom-up approach, the data at the applications planning stage has already been vetted to the business model. If the data has not been examined before the applications are created there is no way to create correct application architecture. Thus even though perfect applications are designed they may not fit the business model. Users will reject the applications if they do not appear to meet their needs or if the numeric results generated by the reports are inaccurate. Without an architecture to reference it is difficult to determine if the inaccuracies are a fault of the applications or of the data.

Plan the technology

Planning the technology requires knowledge of the existing system. System changes will be necessary either to support new products, new technology, system upgrades, acquisitions or new partnerships. For example, many system changes are occurring within the financial services industry in order to comply with new regulations such as Basel II.

It is important to note that systems must continue to function while new technology is being planned to address regulatory needs. Common practice is to apply patches to the existing system rather than revise the overall architecture to support the business model. However, these patches result in changes that affect users upstream and downstream and could lead to data quality issues. If the technology is implemented last, there is more opportunity to phase in necessary changes. Components (subsystems) can be changed and tested with respect to user needs and impacts on data quality can be studied.

Modeling in Practice

As noted by Fitzgerald (1998), no single approach to modeling is actually used in system redesign. A combination of top-down, bottom-up, frameworks etc. is utilized to completely model the system and ensure the required data quality.

In our research, technological fixes were first attempted to meet quality requirements. The IT department, independent yet responsible for the system, was given known errors with the expectation they would correct the errors and apply a patch to prevent the production of future errors. However, the IT department was at the same time charged with developing new applications for new products. Given a finite amount of time and manpower, the data quality is not the top priority.

Data quality was seen as a computer system problem that should be handled by IT. The IT department was having a difficult time supporting the redesign triangle, Figure 3. Data users found accuracy and timeliness problems that incurred a real cost. Though they could identify the cause of the errors and propose corrections they could not change the system. The IT department, fulfilling its charge to integrate new systems and products was stretched to the limit trying to keep up with this continual flow. It was determined that a team effort, cross functional and multilevel was required. The triangle had to be flipped and the business model as well as the system as an enterprise unit had to be studied first. It was concluded that the business processes and governance was not aligned with data quality processes. New regulations forced management to reevaluate their procedures, and alignment. For example, the initial approach did not allow for credit risk exposure validation.

However, the complexity of the system made it difficult to establish line responsibility for system redesign and data quality assurance. All the data feeds containing market prices and volatilities and various credit information data originate from sources external to the credit risk calculating system. By boxing individual systems, such as the Credit Systems Box shown in Figure 5, we were able to determine data quality requirements and business processes that impact credit risk exposure measurement. By reducing the scope of the system being examined, it became possible to define the business model and study the data flow through the boxed area.

When we attempted to measure the data quality in the box, a new organization and system redesign was proposed for the credit risk exposure measurement. And, before this redesign was implemented, yet another new system was proposed based on the "Factory Floor" concept (Snee 2005). System development is an evolutionary and continual process and there is no point in time when the system is fully designed, implemented and static. It is very hard to appraise the quality of information from a system that is continually changing.

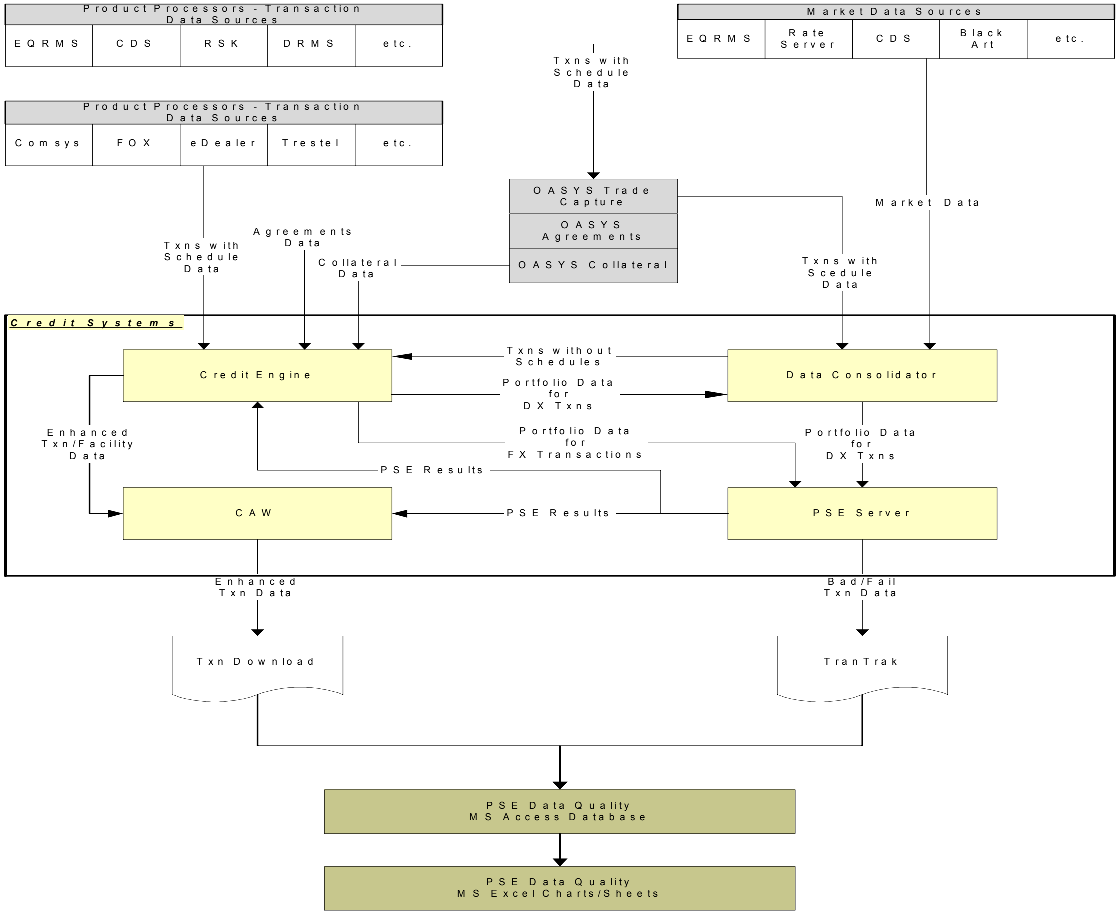

Figure 4
Partial view of system

Since Basel II regulations necessitate validation of credit risk exposure measurement, the continual change in data quality was untenable. Top managers from the technology and administrative areas met to devise a better approach. Yet, they perceived the issues differently. Though each group used the same language and had high quality information as their goal their perspectives were not aligned. The technologist focused on data quality – primarily the existence of a correct value in the field. The managers focused on information quality – reports delivered on time, with all the needed information accurate and correctly displayed. However, these different perspectives did not hamper governance change, since management avoided making premature technological commitments i.e., patches) again. The simple goal to produce the highest quality information for calculating credit risk exposure (the business model) was shared by all and became the point of agreement to which everyone could return.

Conclusions

The reality is never as simple as the model. For a large global enterprise with acquisitions and spin offs, the systems must be flexible and evolve. But, the quality of the information produced by a financial enterprise is paramount to regulatory compliance. A bottom up approach to redesign should improve the sensitivity of the redesign team to both the intrinsic and contextual dimension of data quality, since the business model defines these dimensions.

Data quality should be the joint responsibility of the IT group and business line managers and this team must fully understand the business model. This is the foundation on which data quality should be examined and analyzed and upon which technical decisions should be made.

References

Avison, D., and Fitzgerald, G. Information Systems Development: Methodologies, Techniques and Tools, Blackwell Scientific Publications, Oxford, 1995.

Basel Committee for Banking Supervision, International Convergence of Capital Measurement and Capital Standards: A Revised Framework, Bank for International Settlements, Updated November 2005.

Bubenko,J. "Information System Methodologies-A Research View,". T Olle, H. Sol,
A. Verrijn-Stuart, editors, Information Systems Design Methodologies:Improving the Practice, North-Holland, Amsterdam, 1986, pp 289-318.

Crespi,V., Galstyan,A., and Lerman,K., "Comparative Analysis of Top–Down and Bottom–up Methodologies for Multi–Agent System Design,,",Proceedings of *AAMAS 05* , Utrecht, Netherlands, July 2005, pp1159-1160.

De Amicis ,F., Barone ,D., and Batini, C, "An Analytical Framework To Analyze Dependencies Among Data Quality Dimensions", Proceedings of the Eleventh International Conference on Information Quality , Cambridge, MA, November 2006.

Delez,T., and Hostelttler,D., Proceedings of the Eleventh International Conference on Information Quality , Cambridge, MA, November 2006, pp 235-248.

Earl, M.,"Experiences in Strategic Information Systems Planning **,"** MIS Quarterly, (17:1), Mar., 1993, pp. 1-24.

Fitzgerald, B. "An Empirical Investigation into the Adoption of Systems Development Methodologies," Proceedings of the I International Conference on Information Systems**,** Helsinki, Finland , 1998, pp 103 – 114.

Fleming W., Romas M., Bartlet, D., Rob and Law, A., "Framework for the Effective Adoption of Software Development Methodologies ," ACM SE'06, Melbourne, Florida, USA , March, 10-12, 2006,

Hackos, T., and Redish, C. 1998. User and Task Analysis for Interface Design. New York: Wiley, 1998.

Fisher, C.W. and B.R. Kingman, "Criticality of Data Quality as Exemplified in Two Disasters," Information & Management (39), 2001, pp 109-116.

Fletcher, D., Robbert, M., Middleton, P., and Mohammed, K. "A System Approach to Monitoring Financial Data Quality Assessment and Improvement," The 10th International Conference on Information Quality, Cambridge, MA, November 2005.

Lee, Y.., Strong, D., Kahn, B., and Wang, R.,"AIMQ: A Methodology for Information Quality Assessment," Information and Management, (40:2), 2002, pp 133-146.

Liu,L., and Chi, L., "Evolutional Data Quality: A theory Specific View", Proceedings of the Tenth International Conference on Information Quality, Cambridge, MA, November 2005.

McCorquodale, M., Gebara, F., Kraver, K., Marsman,E., Senger, R., and Richard, B. , "A Top-Down Microsystems Design Methodology and Associated Challenges," Proceedings of the Design, Automation and Test in Europe Conference and Exhibition, 2003, pp1530-591.

Ong, M., ed. The Basel Handbook: A Guide for Financial Practitioners. Risk Books, London, 2005.

Monarchi D. and Puhr G., "A Research Typology for Object-Oriented Analysis and Design", Communications of ACM, (35:.9), Sept. 1992, pp. 35-47.

Raneses, A., Deyong, C., and Glasstein, N., "Modeling Operational Risk in Data Quality", Proceedings of the Eleventh International Conference on Information Quality, Cambridge, MA, November 2006.

Redman, T. C. (1998). “The Impact of Poor Data Quality on the Typical Enterprise.” Communications of the ACM, (41:2), pp. 79-82.

Spewak, S., Enterprise Architecture Planning, WWiley QED: New York, 1992.

Snee, R.,and Hoerl, R., “ Six Sigma Beyond the Factory Floor: Deployment Strategies for Financial Services, Health Care, and the Rest of the Real Economy” , Prentice Hall, 2005

Wand, Y., and Wang, R. ,“ Anchoring Data Quality Dimensions in Ontological Foundations.” Communications of the ACM, (39:11), 1996, pp. 86-95.

Wang, R. , Lee, Y., Pipino, L. , and Strong, D. “ Manage your information as a produc,.”Sloan Management Review, (39:4), 1998, pp. 95-105.

Woo,F., Mikusauskas,R., Dean Bartlett, and Law, R,, “A Framework for the Effective Adoption of Software Development Methodologies ,”ACM Software Engineering 06, Melbourne, Florida, March, 10-12, 2006.

Yonke, l.,,Palermo, R., and Bates, R., “The Road to Data Quality at Aera Energy LLC”, Panel at ICIQ, Cambridge, MA, 2006.